

False starts

Net prophets Sydney Brenner



Statements that “we have come to do biology in a new way” or “there is a new paradigm in biological research” are now commonplace. Nobody seems to be satisfied by a single good experiment that gives a precise answer to a well formulated

question, which was the old way we did biology. On the contrary there is now a belief that a mass attack on parallel fronts can provide a database of all the information in one concerted effort, and all we need is a computer programme that will give everybody all the knowledge they need.

Much of this stems from genome projects, especially the effort to sequence the human genome. However, there are subtle differences between the different cultures that have generated the sequences. The yeast genome was sequenced by a co-operative venture of many small individual scientific groups, who had a deep interest in the result. Surrounding the project was an even larger group of yeast geneticists and molecular biologists who knew how to use the sequence in their experimental work. The sequence was the path to the genes of yeast; there are now ways to access all of the genes directly and the page in the Book of Life devoted to yeast is written in real DNA. The sequence has become the tool for research that it was expected to be, and not an end in itself.

It is likely that the genome projects for *Caenorhabditis elegans* and *Drosophila* will have the same impact on their fields, mainly because of the large number of researchers who can

immediately make use of the product. It is with the vertebrate genomes that we find a new idea coming to the fore. Roughly speaking, the proponents have come to believe that computers can extract biological significance directly from DNA sequences.

This approach has generated two new areas of activity. One, Bioinformatics, is simply pretentious; the other, Functional Genomics, is ridiculous. The latter uses the former to try to find function from the sequences of genes. I don't think that there are any university departments devoted to these subjects but there are certainly a growing number of companies doing one or both. Other areas are now adopting the same approach of systematically assembling data by factory methods. The proteome is emerging from two-dimensional electrophoresis of proteins, but is still a poor relation of the genome. I expect to see the glycome and the lipome next.

Actually, there is already a perfectly good name for the science of studying gene function; it used to be called Genetics. Geneticists have always been interested in function and have always used their research as a way — perhaps the way — to analyse complex functions of organisms. The sequences of genes and, better still, the pieces of DNA that correspond to the genes, replace what could only be achieved by the mutant hunt in classical experimental genetics; they are tools and not ends in themselves. We will still need to find out how each gene works and piece together the elaborate network of gene interactions by the old paradigm of experiment. In fact, sequences also offer us the possibility of interpreting Nature's experiments in evolution, but that will come later as a consequence of knowing the genetics of contemporary organisms.

Bioinformatics has its place. Its main activity has been beneficial in that masses of data can now be easily reached and used for research. However, the idea that sequence data can have other information added to

them which will give us knowledge of function is surely misplaced. For this, we must do more than repackage what is known; the computers must compute, and in order to do this we need a theory that we can test. The subject that will be developed will be one that should be called Theoretical Biology, but as this has a bad name we call it Computational Biology.

The siliconization of biology has been successful — perhaps too successful — in one area, which is in the way we communicate. I note that many researchers are now spending several hours a day with their e-mail, reading and sending messages to an increasing number of correspondents. I fear that this is going to put everybody in an electronic committee in permanent session. I have installed a very narrow pore filter on my e-mail; I have someone else read it and print out what I need to know. I started this mainly because a dentist in Philadelphia sent me voluminous messages about his new theories on the brain, and also because I cannot remember my password.

More than ten years ago, when electronic mail was still a novelty, I was given an account on a private network. Three passwords were requested to enter the system, and had to be renewed at frequent intervals for reasons of security. I used all twenty amino acids and the five nucleotide bases, and I then started on them again but written backwards, which makes a surprising list from which I particularly liked enilav, but there is also a enicuelosi, which has a good Italian ring to it. At the risk of compromising my computer security I shall disclose my favourite password which is ELCID, usually with some number attached because greedy computers want six characters. This password lets me login to the computer but apparently another one is needed for e-mail, which is a secret even from me. I am also toying with the idea of having a special address for bioinformaticists and functional geneticists to reach me. How about unclesyd@gnome.zurich.pri ?